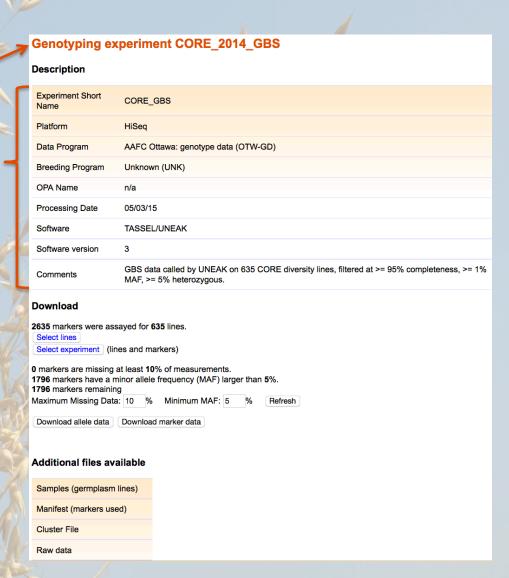# T3/Oat tutorials

**Tutorial 6:** Uploading genotype experiment descriptions and data

https://triticeaetoolbox.org/oat/

# T3 genotype trial reports

- The experiment name is the unique identifier for a T3 genotype trial

- The experiment description is loaded first

- A line translation file must then be loaded, which links existing T3 lines to the genotype experiment

- Genotype data can then be uploaded in one of four formats

- Supplemental data files can also be attached to the genotype experiment

**Genotyping experiment CORE_2014_GBS**

**Description**

| Experiment Short Name | CORE_GBS |
|---|---|
| Platform | HiSeq |
| Data Program | AAFC Ottawa: genotype data (OTW-GD) |
| Breeding Program | Unknown (UNK) |
| OPA Name | n/a |
| Processing Date | 05/03/15 |
| Software | TASSEL/UNEAK |
| Software version | 3 |
| Comments | GBS data called by UNEAK on 635 CORE diversity lines, filtered at >= 95% completeness, >= 1% MAF, >= 5% heterozygous. |

**Download**

**2635** markers were assayed for **635** lines.
Select lines
Select experiment (lines and markers)

**0** markers are missing at least **10%** of measurements.
**1796** markers have a minor allele frequency (MAF) larger than **5%**.
**1796** markers remaining
Maximum Missing Data: 10 %    Minimum MAF: 5 %    Refresh

Download allele data    Download marker data

**Additional files available**

| Samples (germplasm lines) |
|---|
| Manifest (markers used) |
| Cluster File |
| Raw data |

# Tutorial 6: Outline

1. Downloading the data submission templates
2. The experiment description template
   1. Completing the experiment description template
   2. Test loading the experiment description template
3. Completing the line translation template
4. Completing the genotype data template
5. Test loading genotype results
6. Submitting the completed templates for upload to T3/Oat

# Section 1: Downloading the data submission templates

Choose the About T3 menu> Data Submission

# Section 1: Downloading the data submission templates

| Genotyping | | | |
|---|---|---|---|
| | Experiment description | 30Sep13 | Platform, software, manifest file, experiment details... |
| | Line translation | 04Nov11 | Line Name and Trial Code |
| | Illumina data | 03Jan13 | alleles for lines and markers (coded as A or B) |
| | GBS data | 11Apr14 | alleles for lines and markers (coded as ACTG, N = missing, H = heterozygous) |
| | GBS data, over 100K markers | 07Nov14 | instructions for loading large GBS data sets using the command line |
| | DArT data | 17Jul13 | alleles for lines and markers (coded as 1 = present, 0 = absent) |
| | Experiment results (1D) | 16Sep11 | 1D table of alleles for lines and markers |

The "Genotyping" section of the "Data Submission" page contains the three templates that you need to load genotype data:

1. The **"Experiment description"** template is used to submit trial details. The template can be used to create a new trial record or to update existing genotype trial information in T3/Oat.

2. The **"Line translation"** template links the T3 lines that were genotyped in the trial to the genotype trial name.

3. The genotype data is loaded using either the **"Illumina data"**, **"GBS data"** or **"DArT data"** template, or as a 1D table of alleles.

# Section 1: Downloading the data submission templates

- Each of the "Genotyping" template links will take you to a webpage displaying the template in .txt format
- Right click the link and "Save Link As…" to store the text file to your computer
- This tutorial will use Excel to display the text files

```
Breeding Program        CAPdata Program Year     Short Name      Trial Code       Traits
Platform        Processing Date Manifest File   Cluster File   OPA Name         Analysis
Software        Software Version        Sample Sheet    Comments
AB      NDG     2010    AB_BOPA1        2009BOPA1_AB_Plate5      SNP calls, r, theta, x,
y,xraw, yraw    GoldenGate      04/28/2010      GS0007511-OPA.opa
BOPA1PooledSamples2009.EGT      BOPA1   Illumina Genome Studio   2010.3
Sample_Sheet_2009plate5AB.csv
AB      NDG     2010    AB_BOPA2        2009BOPA2_AB_Plate5      SNP calls, r, theta, x,
y,xraw, yraw    GoldenGate      04/28/2010      GS0010598-OPA.opa
BOPA2PooledSamples2009.EGT      BOPA2   Illumina Genome Studio   2010.3
2009CAPplate5ABSampleSheetBOPA2rev.csv
MN      NDG     2010    MN_BOPA1        2009BOPA1_MN_Plate3      SNP calls, r, theta, x,
y,xraw, yraw    GoldenGate      4/16/2010       GS0007511-OPA.opa
BOPA1PooledSamples2009.EGT      BOPA1   Illumina Genome Studio   2010.3
Sample_Sheet_2009plate3MN.csv
------------------- cut
----------------------


ADDITION
INFORMATION


These are tab-delimited files of the
form:

Breeding_Program        VT (Two letter
code)
CAPdata_Program NDG (Should be the same as long as North Dakota genotyping is doing
this)
Year
2008
Short_Name      VT_Plate1 (Plate number stays the same year to
year)
Trial_code      2008BOPA1_VT_Plate1  (This is the basic code, CAPYEAR, BOPA
#_CAPCODE_Plate#)

Traits  SNP calls, r, theta, x, y,xraw,
yraw
Platform        GoldenGate
Processing_date
8/9/2008

manifest_file   GS0007511-OPA.opa  (need to set for each
BOPA)
cluster_file    BOPA1PooledSamples2009.EGT  (need to set for each
BOPA)
OPA_name        BOPA1  (need to set for each
BOPA)
Analysis_Software       Illumina Genome
Studio
Software_Version
2010.3
Sample_sheet    Sample_Sheet_2008plate1-10BOPA1.csv  (need to set for each Sample
Sheet)
```

# Section 2.1: Completing the experiment description template

The notes section contains nearly all of the information needed to fill out the experiment description template, apart from:

– Breeding and data program codes

This information can be found under the "About T3" menu.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Breeding Program | CAPdata Program | Year | Short Name | Trial Code | Traits | Platform | Processing Date | Manifest | Cluster File | OPA Name | Analysis Software |
| 2 | UNK | OTW-GD | 2014 | CORE_GBS | CORE_2014_GBS | GBS SNP calls (A/C/G/T) | HiSeq | 5/3/15 | n/a | n/a | n/a | TASSEL/UNEAK |
| 3 | UNK | FND-GD | 2010 | CORE_Infinium | CORE_2010_Infinium | SNP calls | Infinium | 7/11/14 | n/a | CORE_oat6Kproject.bsc | Oat6K | Illumina Genome Studio; Illuminizer |
| 4 | -------------------- cut ---------------------- | | | | | | | | | | | |
| 5 | | | | | | | | | | | | |
| 6 | ADDITIONAL INFORMATION | | | | | | | | | | | |
| 7 | | | | | | | | | | | | |
| 8 | These are tab-delimited files of the form: | | | | | | | | | | | |
| 9 | Rows can be inserted above the "cut" row in order to submit more genotype experiments in one document. Please remove rows that are not in use. | | | | | | | | | | | |
| 10 | Breeding_Program | The 3-letter breeding program code can be found under the About T3 menu> Contributing Data Programs. Use the unknown code "UNK" when multiple breeding programs contributed lines for the trial. | | | | | | | | | | |
| 11 | CAPdata_Program | The 5-letter data program code for the genotyping lab can be found under the About T3 menu> Contributing Data Programs. | | | | | | | | | | |
| 12 | Year | The year that the lines were genotyped. | | | | | | | | | | |
| 13 | Short_Name | The Trial_code (below) minus one of the three elements of the name (at submitters discretion). | | | | | | | | | | |
| 14 | Trial_code | The unique identifier for a genotype trial, usually in the format: ProjectName_YYYY_Platform/Some other identifying feature of the trial. The trial code should not contain any spaces. | | | | | | | | | | |
| 15 | Traits | Unrestricted field, typically "SNP calls" or "GBS SNP calls (A/C/G/T)". | | | | | | | | | | |
| 16 | Platform | "Infinium", "HiSeq", "Golden Gate OPA" or "GBS Pstl-Mspl". Please contact the curator if you require an additional platform to be added to T3/Oat. | | | | | | | | | | |
| 17 | Processing_date | M/D/YY (month and day without preceeding zeros) | | | | | | | | | | |
| 18 | manifest_file | The file name of an optional manifest file to be uploaded in addition to the genotype results. | | | | | | | | | | |
| 19 | cluster_file | The file name of an optional cluster file to be uploaded in addition to the genotype results. | | | | | | | | | | |
| 20 | OPA_name | OPA name | | | | | | | | | | |
| 21 | Analysis_Software | The software that was used for analysis. | | | | | | | | | | |
| 22 | Software_Version | Version of the software that was used for analysis. | | | | | | | | | | |
| 23 | Sample_sheet | The file name of an optional sample sheet to be uploaded in addition to the genotype results. | | | | | | | | | | |
| 24 | Comments | Provide necessary additional information that cannot be entered in any other field (optional). | | | | | | | | | | |

# Section 2.1: Completing the experiment description template

To find the breeding and data program codes
- Choose the About T3 menu> Contributing Data Programs

# Section 2.1: Completing the experiment description template

- Both the 3-letter breeding program code and the 5-letter data program code of the genotyping lab can be found on this page
- Contact the curator if you do not find an appropriate code

## Contributing Programs

**Breeding** programs contribute lines.

| Breeding Program | Code | Collaborator | Description | Institution |
|---|---|---|---|---|
| AAES, Auburn University | AUB | Kathryn Glass | Alabama Agricultural Experiment Station (AAES), Auburn University, AL-USA. | Auburn University |
| AAFC Agassiz | ABC | | Agriculture and Agri-Food Canada (AAFC) Pacific Agri-Food Research Centre (PARC) in Agassiz, BC-CAN. | Agriculture and Agri-Food Canada |
| AAFC Brandon | MTB | Jennifer W. Mitchell-Fetch | Agriculture and Agri-Food Canada (AAFC) Brandon Research Centre, MB-CAN. | Agriculture and Agri-Food Canada |
| AAFC Lacombe | LAC | Jennifer W. Mitchell-Fetch | Agriculture and Agri-Food Canada (AAFC) Lacombe Research Centre, AB-CAN. | Agriculture and Agri-Food Canada |
| AAFC Ottawa | OTW | Weikai Yan | Agriculture and Agri-Food Canada (AAFC) Eastern Cereal and Oilseed Research Centre (ECORC) in Ottawa, ON-CAN. | Agriculture and Agri-Food Canada |
| AAFC Sainte-Foy | STE | | Agriculture and Agri-Food Canada (AAFC) Soils and Crops Research and Development Centre (SCRDC) in Quebec | Agriculture and Agri-Food |

# Section 2.2: Test loading the experiment description template

- Navigate to the Oat Sandbox at https://t3sandbox.org/t3/sandbox/oat/
- The curation menu will appear once you register and login
- Choose the Curate menu> Genotype Experiments

# Section 2.2: Test loading the experiment description template

The sandbox allows you to test load your template. The Oat Sandbox reverts to an exact copy of T3/Oat overnight.



**Upload Genotype Experiment Annotation**

**Genotype Annotation File:** [ Browse... ]  No file selected.
Example Genotype Annotation File

Do you want this data to be Public?  ● Yes    ○ No
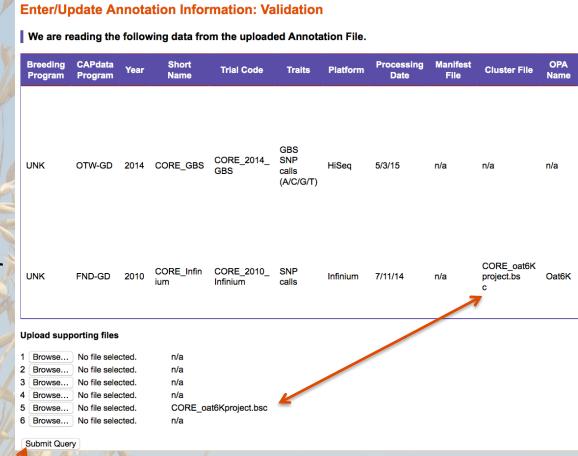[ Upload Annotation File ]

It is possible to load "private data" to the main T3/Oat site. Private data is only visible to T3/Oat users that are designated as collaborators by the site administrators. Data may remain private for a limited time only. Please contact the curator if you are interested in this option.

# Section 2.2: Test loading the experiment description template

The validation window shows how T3 has read the upload document
- Please take the time to validate your data

Attach any available supporting documents
- Remember to attach a manifest file, cluster file or sample sheet if you referred to them in the experiment description
- Additional supporting files can be loaded at your discretion

**Enter/Update Annotation Information: Validation**

We are reading the following data from the uploaded Annotation File.

| Breeding Program | CAPdata Program | Year | Short Name | Trial Code | Traits | Platform | Processing Date | Manifest File | Cluster File | OPA Name |
|---|---|---|---|---|---|---|---|---|---|---|
| UNK | OTW-GD | 2014 | CORE_GBS | CORE_2014_GBS | GBS SNP calls (A/C/G/T) | HiSeq | 5/3/15 | n/a | n/a | n/a |
| UNK | FND-GD | 2010 | CORE_Infinium | CORE_2010_Infinium | SNP calls | Infinium | 7/11/14 | n/a | CORE_oat6K project.bsc | Oat6K |

**Upload supporting files**

1  Browse…  No file selected.  n/a
2  Browse…  No file selected.  n/a
3  Browse…  No file selected.  n/a
4  Browse…  No file selected.  n/a
5  Browse…  No file selected.  CORE_oat6Kproject.bsc
6  Browse…  No file selected.  n/a

Submit Query

Don't forget to "Submit" the upload if the information is correct.

# Section 2.2: Test loading the experiment description template

```
no file upload for entry 1
no file upload for entry 2
no file upload for entry 3
no file upload for entry 4
no file upload for entry 5
no file upload for entry 6
The Manifest file is missing, n/a
The Cluster file is missing, n/a
The Sample Sheet file has is missing, n/a

Error: Breeding program 'XYZ' is not in the database.

Return
```

```
no file upload for entry 1
no file upload for entry 2
no file upload for entry 3
no file upload for entry 4
no file upload for entry 5
no file upload for entry 6
The Manifest file is missing, n/a
The Cluster file is missing, n/a
The Sample Sheet file has is missing, n/a
update experiment CORE_2014_GBS
The Manifest file is missing, n/a
The Cluster file is missing, CORE_oat6Kproject.bsc
The Sample Sheet file has is missing, n/a
update experiment CORE_2010_Infinium
The Data is inserted/updated successfully


Go Back To Main Page
```

- It is OK to have a "missing" manifest file, cluster file or sample sheet if those files are unavailable to you
- Any illegal entries will be highlighted at this stage
  e.g. invalid breeding program code
- Fix illegal entries in your upload document and re-submit it

Once your experiment description has loaded successfully you can upload:
- The line translation file
- The genotype data file

These files are uploaded at the same time

# Section 4: Completing the line translation template

| | A | B |
|---|---|---|
| 1 | Line Name | Trial Code |
| 2 | AAC_BULLET | CORE_2014_GBS |
| 3 | AAC_ROSKENS | CORE_2014_GBS |
| 4 | AARRE | CORE_2014_GBS |
| 5 | AC_MARIE | CORE_2014_GBS |
| 6 | AC_MORGAN | CORE_2014_GBS |
| 7 | AC_RIGODON | CORE_2014_GBS |
| 8 | AJAX\|CORE | CORE_2014_GBS |
| 9 | AJAY | CORE_2014_GBS |
| 10 | AKIYUTAKA | CORE_2014_GBS |
| 11 | ALLEN | CORE_2014_GBS |
| 12 | ANDREW | CORE_2014_GBS |
| 13 | ARDENTE | CORE_2014_GBS |
| 14 | ASENCAO | CORE_2014_GBS |
| 15 | ASSINIBOIA_S42 | CORE_2014_GBS |
| 16 | AURORA | CORE_2014_GBS |
| 17 | AVE117 | CORE_2014_GBS |
| 18 | AVENY | CORE_2014_GBS |
| 19 | BALADO | CORE_2014_GBS |
| 20 | BARRA | CORE_2014_GBS |
| 21 | BELINDA | CORE_2014_GBS |
| 22 | BETAGENE | CORE_2014_GBS |
| 23 | BIA | CORE_2014_GBS |
| 24 | BIRI | CORE_2014_GBS |
| 25 | BLAZE | CORE_2014_GBS |

The line translation file links existing T3 line records to the genotype trial.

**The first column** should contain the names of genotyped lines as they appear in T3/Oat.
**The second column** should contain the trial code that was created in the experiment description upload document, repeated for every line.

Please refer to T3/Oat tutorial 1 for instructions on how to:

- Correctly format line names for T3/Oat
- Search for existing line records by line name, using the "Select Lines by Properties" tool
- Add new line records to T3/Oat

# Section 5: Completing the genotype data template

**Illumina data template**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | SNP | AAC_BULLET | AAC_ROSKEN | AC_MARIE | AC_MORGAN |
| 2 | BA_grs_c103 | BB | BB | AA | AA |
| 3 | GMI_DS_A3_ | AA | AA | AA | AA |
| 4 | GMI_DS_A3_ | BB | BB | BB | BB |
| 5 | GMI_DS_A3_ | AA | AA | BB | BB |
| 6 | GMI_DS_A3_ | AA | AA | AA | AA |
| 7 | GMI_DS_A3_ | BB | BB | BB | BB |
| 8 | GMI_DS_A3_ | BB | AA | BB | BB |
| 9 | GMI_DS_A3_ | BB | BB | BB | BB |
| 10 | GMI_DS_A3_ | AA | AA | AA | AA |

**GBS data template**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | SNP | AAC_BULLET | AAC_ROSKEN | AC_MARIE | AC_MORGAN |
| 2 | avgbs_10001 | T | T | T | T |
| 3 | avgbs_10005 | T | N | T | T |
| 4 | avgbs_10009 | T | T | T | T |
| 5 | avgbs_10026 | T | T | T | T |
| 6 | avgbs_10045 | T | T | T | T |
| 7 | avgbs_10059 | G | G | N | G |
| 8 | avgbs_10071 | C | C | C | C |
| 9 | avgbs_10084 | A | C | A | C |
| 10 | avgbs_10086 | G | G | G | G |

**DArT data template**

| | A | B | C |
|---|---|---|---|
| 1 | SNP | DAVEM | CAYUGAX |
| 2 | bPb-2121 | 1 | 0 |
| 3 | bPb-7983 | 0 | 1 |

Genotype data can be loaded as:

1. Illumina data
   - Alleles coded as A or B
2. GBS data
   - Alleles coded ACTG
   - N= missing
   - H= heterozygous
3. DArT data
   - Alleles coded as 1 (present) or 0 (absent)
4. 1D table of alleles for lines and markers (not shown)

In cases 1-3, the first column should contain the T3 marker names and the first row should contain the T3 line names.

# Section 5: Completing the genotype data template

**Illumina data template**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | SNP | AAC_BULLET | AAC_ROSKEN | AC_MARIE | AC_MORGAN |
| 2 | BA_grs_c103 | BB | BB | AA | AA |
| 3 | GMI_DS_A3_ | AA | AA | AA | AA |
| 4 | GMI_DS_A3_ | BB | BB | BB | BB |
| 5 | GMI_DS_A3_ | AA | AA | BB | BB |
| 6 | GMI_DS_A3_ | AA | AA | AA | AA |
| 7 | GMI_DS_A3_ | BB | BB | BB | BB |
| 8 | GMI_DS_A3_ | BB | AA | BB | BB |
| 9 | GMI_DS_A3_ | BB | BB | BB | BB |
| 10 | GMI_DS_A3_ | AA | AA | AA | AA |

**GBS data template**

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | SNP | AAC_BULLET | AAC_ROSKEN | AC_MARIE | AC_MORGAN |
| 2 | avgbs_10001 | T | T | T | T |
| 3 | avgbs_10005 | T | N | T | T |
| 4 | avgbs_10009 | T | T | T | T |
| 5 | avgbs_10026 | T | T | T | T |
| 6 | avgbs_10045 | T | T | T | T |
| 7 | avgbs_10059 | G | G | N | G |
| 8 | avgbs_10071 | C | C | C | C |
| 9 | avgbs_10084 | A | C | A | C |
| 10 | avgbs_10086 | G | G | G | G |

**DArT data template**

| | A | B | C |
|---|---|---|---|
| 1 | SNP | DAVEM | CAYUGAX |
| 2 | bPb-2121 | 1 | 0 |
| 3 | bPb-7983 | 0 | 1 |

- The line names used in the genotype data submission document must match the line names in the related line translation document
- All lines and markers in the submission documents must already be present in T3/Oat

Please refer to T3/Oat tutorial 4 for instructions on how to:

- Search for existing marker records using the "Select Markers" tool
- Add new marker records to T3/Oat

# Section 6: Test loading genotype results

- Navigate to the Oat Sandbox at https://t3sandbox.org/t3/sandbox/oat/
- The curation menu will appear once you register and login
- Choose the Curate menu> Genotype Results

# Section 6: Test loading genotype results

**Add Genotype Experiment Information**

**Line Translation File:** [Browse...] No file selected.

Example Line Translation File

**Genotype Data File:** [Browse...] No file selected.    Size limit 100M

**Data File Format:**
- ○ 1D Example Genotype Data File
- ● 2D Example Illumina_Genotype_template.txt
- ○ 2D Example GBS_Genotype_template.txt (ACTG, N = missing, H = heterozygous)
- ○ 2D Example DArT_Genotype_template.txt (1 = Present, 0 = Absent, "-" = missing)

[Upload Line Translation and Genotype Data File]

[Verify Files] Check if lines and markers exist in database

Note:
1. Both files (line translation and genotype data) are required. Line names should be identical between both files.
2. Due to size of the Genotype Data File, it can be compressed with a "zip" application before submitting it.
3. Genotype files with over 100K makers should be imported via the command line as described in GBS import instruction.
4. This upload process may take several hours to complete depending on size of the data file. Please leave your email address for us to contact you with the results.

**Your Email Address:** [                    ]

Attach your completed line translation file…

Attach your results file and select which of the genotype data templates you used…

You'll receive an email when your upload was successful (or an email telling you why it was not)

# Section 7: Submitting the completed templates to T3/Oat

Finally, submit your documents to the curator for upload to T3/Oat

- Navigate to the Data Submission page
- Click the "Submit" button

Contact Us

## T3/Oat Sandbox

Home   Select ▾   Analyze ▾   Download ▾   Curate ▾   About T3 ▾

**Quick Links**

**Current selections:**

Lines: 0

Markers: All

Traits: 0

Phenotype Trials

Genotype Experiments

Quick search...

**What's New**

*February 2016*

**Genetic Maps**

### Data Submission

- Data templates are Excel, .txt, or .csv worksheets with column headers for the data T3 requires or accepts.
- The example values can be replaced with your own. Notes about the restrictions for valid data are included in the templates.
- Once populated, the files can be loaded into T3 using the **Curate** menu, which is available to registered Sandbox users.
- To make updates or corrections, edit your file and reload.
- Please use the oat Sandbox database for test-loading your files. When they're ready, click here to submit them to the T3 Curator for loading into the official database.

Submit

**Instructions** - Rules for filling in the templates, and sequence of submission

**Tutorials**

- Lesson One. Germplasm file creation and upload
- Lesson Two. (.pdf) Phenotype trial descriptions and data
- Lesson Three. Genotype trial descriptions and data
- Lesson Four. How to create germplasm line panels in T3

# Section 7: Submitting the completed templates to T3/Oat



You can submit files that do not successfully upload to the Oat Sandbox to receive help from the curator

# Contact Us

Please contact us if you need help using T3/Oat